



## EVOLUÇÃO ESPAÇO-TEMPORAL DO CAFÉ NO ESTADO DE MINAS GERAIS POR MEIO DO CLASSIFICADOR ÁRVORE DE DECISÃO

Rafael Coll Delgado<sup>1</sup>, Gilberto Chohaku Sedyama<sup>2</sup>, Evaldo de Paiva Lima<sup>3</sup>, Evandro Chaves de Oliveira<sup>4</sup>, Rafael de Ávila Rodrigues<sup>5</sup>

1. Professor Doutor da Universidade Federal Rural do Rio de Janeiro, Seropédica/RJ – Brasil ([rafaelcolldelgado32@gmail.com](mailto:rafaelcolldelgado32@gmail.com))
2. Professor Ph. D. da Universidade Federal de Viçosa, Viçosa/MG – Brasil
3. Pesquisador Doutor da Embrapa Solos, Rio de Janeiro/RJ – Brasil
4. Professor Doutor do Instituto Federal de Educação, Ciência e Tecnologia do Espírito Santo, IFES – Campus Itapina – Brasil
5. Pesquisador Doutor da Universidade Federal de Viçosa, Viçosa/MG – Brasil

Recebido em: 06/10/2012 – Aprovado em: 15/11/2012 – Publicado em: 30/11/2012

### RESUMO

Neste trabalho realizou-se a classificação espectral da área plantada de café por meio do classificador árvore de decisão (AD), onde analisou-se a influência do clima e relevo. Para a classificação espectral das áreas plantadas, empregou-se o algoritmo AD e produtos do sensor TM, a bordo do satélite Landsat-5. A classificação espectral por meio do algoritmo AD mostrou que, as imagens de sensoriamento remoto obtida pelo satélite Landsat-5, em anos diferentes e épocas distintas, permitem mapear as áreas ocupadas com a cultura do café no Estado de Minas Gerais. Os valores das áreas plantadas, estimadas pelo algoritmo AD, apresentaram boa correlação com os valores observados pelo Instituto Brasileiro de Geografia e Estatística (IBGE), porém houve uma tendência de superestimativa para os anos de 1995 e 2000 com Viés Médio (VM) de 4596,96 e 13689,14 ha, respectivamente. Entretanto, nos anos de 1990 e 2005 ocorreu subestimativa dos resultados de área plantada, com valores, respectivamente, de -255,51 e -4433,39 ha. Ao comparar as áreas plantadas estimadas via AD com os dados do IBGE, verificou-se que os valores de AD foram compatíveis com a literatura.

**PALAVRAS CHAVE:** Classificação espectral, área plantada, Landsat 5.

### SPACE-TIME EVOLUTION OF COFFEE IN THE STATE OF MINAS GERAIS THROUGH THE DECISION TREE CLASSIFIER

### ABSTRACT

This work aimed to carry out the spectral classification of the area cultivated with coffee through the decision tree classifier (DT), in where there have been analyzed the influence of climate and topography. For the spectral classification of cultivated areas, it was used the algorithm DT and products from TM sensor aboard the Landsat-5. The spectral classification by the DT algorithm showed that the images obtained by remote sensing through satellite Landsat-5 in four different years and

specific times, allow the mapping of the areas occupied by the coffee crop at the state of Minas Gerais. The values of the cultivated areas, estimated by the algorithm DT, showed good correlation with the values observed by the Brazilian Institute of Geography and Statistics (*Instituto Brasileiro de Geografia e Estatística* - IBGE), but there was a tendency of overestimation for the years 1995 and 2000 with Bias Medium (BM) OF 4596.96 ha and 13689.14 ha, respectively. However, there has been underestimation of the results of area cultivated with coffee for the years 1990 and 2005, with respective values of -255.51 ha and -4433.39 ha. By comparing the DT's estimated cultivated areas to IBGE data, there was consistency in the results based on the literature.

**KEYWORDS:** Spectral classification, acreage, Landsat 5.

## INTRODUÇÃO

O levantamento da área agrícola é uma informação fundamental no planejamento agrícola e no planejamento do território como um todo, seja na questão econômica, agrária, ambiental, ou social. A extensão da área agrícola, além de ser uma componente no cálculo da produção agrícola do território, é, por exemplo, uma variável no cálculo de impostos. No Brasil, as estimativas oficiais de produtividade agrícola e área plantada são realizadas predominantemente de forma subjetiva, por meio de entrevistas aplicadas aos produtores, informações sobre a venda de insumos, análise de safras passadas, observação das condições meteorológicas, entre outros, o que não permite uma análise apurada dos erros envolvidos e dá margem a especulações (RIZZI et al., 2007).

A amostragem tem sido a maneira mais utilizada pelo IBGE (Instituto Brasileiro de Geografia e Estatística) e pela CONAB (Companhia Brasileira de Abastecimento) para a estimativa nacional de área agrícola, sendo a mais utilizada a forma não probabilística. Contudo, a abordagem amostral probabilística pode ser uma alternativa para tais estimativas, pois oferece possibilidade de quantificação de erros e variâncias associadas, além de atender à crescente demanda por resultados menos subjetivos (EPIPHANIO et al., 2002; ADAMI et al., 2004). Neste sentido, o uso do Sensoriamento Remoto (SR) e Sistema de Informação Geográfica (SIG) permitem a aplicação de técnicas probabilísticas, amostrais e de mapeamento, e podem oferecer maior operacionalidade e menor custo de execução para as estimativas de área (GONZÁLES-ALONSO et al., 1991; FAO, 1996).

Existem diferentes técnicas de reconhecimento de padrões em imagens de SR que podem auxiliar na identificação de áreas com maior probabilidade de ocorrência e auxiliar em estratificações espaciais, a mineração de dados por árvore de decisão, apesar de ser uma técnica relativamente nova, é relatada como promissora na análise de dados de SR.

DELGADO et al. (2012) mapearam e classificaram espectralmente a cultura da cana-de-açúcar em diferentes épocas de plantio na Fazenda Boa Fé, localizada no Triângulo Mineiro, no município de Conquista, Minas Gerais, e concluíram que o classificador árvore de decisão apresentou potencial para o mapeamento das áreas cultivadas com cana-de-açúcar.

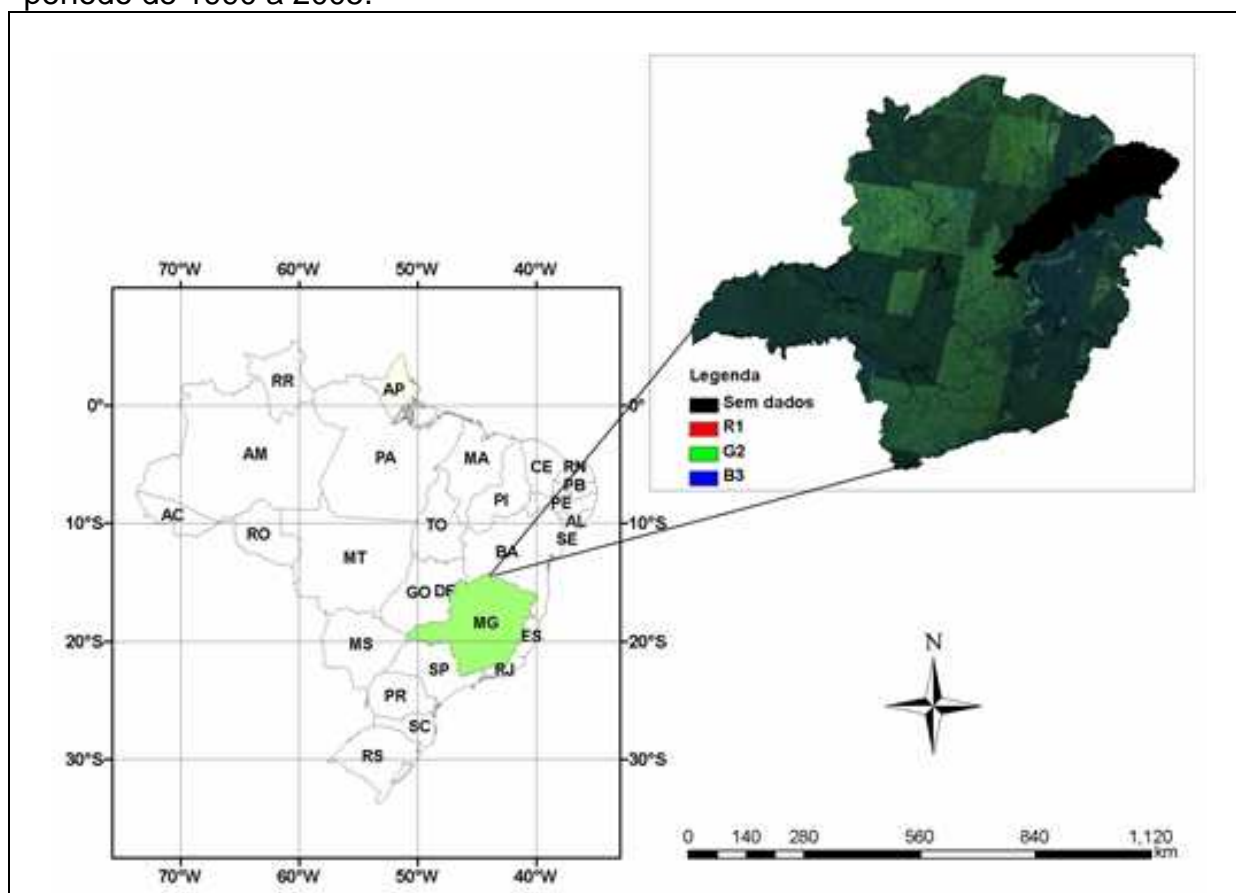
CARVALHO et al. (2008) mapearam a Floresta de Mata Atlântica na Serra dos Órgãos-RJ, com o algoritmo árvore de decisão para que pudessem integrar dados de Sensoriamento Remoto e modelo digital de elevação (MDE), utilizando para esse estudo imagens do sensor ASTER/TERRA e fotografias aéreas da região. A

imagem ASTER foi primeiramente classificada e tratada com a ajuda da fotografia e, posteriormente, essa classificação foi usada como informação junto com o MDE para a árvore de decisão. As classes delimitadas foram: Agropecuária, Campos de Altitude, Floresta Ombrófila Densa Alto-Montanha, Floresta Ombrófila Densa Montana, Floresta Ombrófila Densa Submontana, Floresta Secundária e áreas sem vegetação. A imagem distinguiu ambientes florestais e não florestais, e o MDE distinguiu áreas de campos. Os autores afirmaram que a árvore de decisão foi extremamente eficiente para a classificação de dados oriundos de diferentes sensores.

Considerando a importância do café na economia mineira, este estudo tem o objetivo de avaliar a adequação do algoritmo classificador árvore de decisão na discriminação e mapeamento do uso do solo pela cultura do café por meio de imagens do satélite Landsat-5.

## METODOLOGIA

A área de estudo compreende o Estado de Minas, situado na Região Sudeste do Brasil, entre os paralelos 14°13'57" e 22°55'22" de latitude Sul e os meridianos 39°51'23" e 51°02'45" de longitude Oeste de Greenwich (Figura 1). Na Figura 1 a mesorregião Jequitinhonha não foi considerada na classificação de área plantada por árvore de decisão, devido à influência de nuvens e ausência de imagens para o período de 1990 a 2005.

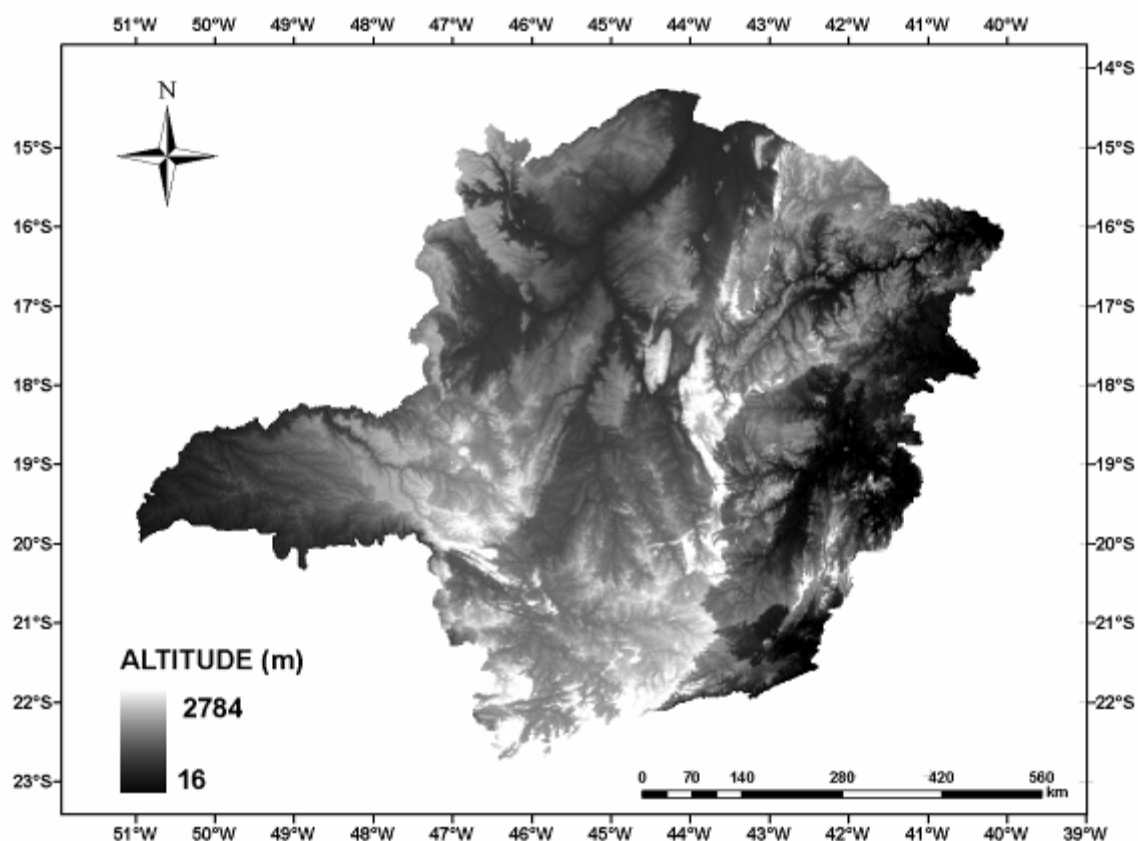


**FIGURA 1.** Localização geográfica da área de estudo e, no segundo plano, mosaico da imagem do sensor TM com a composição RGB (1, 2 e 3).

Fonte: elaborado por Rafael Coll Delgado, julho de 2012.

Os dados de imageamento orbital utilizados para determinar as áreas de café compreendem imagens do sensor TM a bordo do satélite Landsat-5 dos anos de 1990, 1995, 2000 e 2005. As imagens Landsat-5 são em grande maioria do período seco, em que se verifica pouca cobertura de nuvens e também foram utilizadas das imagens tomadas em outras estações dos anos em estudo para aferição de eventuais dúvidas quanto à caracterização da cultura do café em cenários dinâmicos, como também para prover uma base de dados histórica sobre a área plantada existente desta cultura na área de estudo.

Quanto ao relevo, utilizou-se dados gerados a partir de dados de radar, obtidos de sensores a bordo do ônibus espacial Endeavour, no projeto SRTM (*Shuttle Radar Topography Mission*) (Figura 2). Como suporte à interpretação dessas imagens, foram utilizados dados de censo agrícola de 1990 a 2005, disponíveis no site do Instituto Brasileiro de Geografia e Estatística, referente a área plantada das culturas estudadas neste trabalho. Para o processamento dos dados foram utilizados os softwares ENVI© (RSI) versão 4.3, ERDAS IMAGINE 9.1 e ArcGIS 9.2.



**FIGURA 2.** Modelo Digital de Elevação (MDE) para o Estado de Minas Gerais.

Na classificação pelo algoritmo de classificação por árvore de decisão, foi utilizada a metodologia conforme proposto por DELGADO et al. (2012), e um aplicativo de processamento de imagens, para a classificação digital. Nesta técnica de classificação, foram utilizadas imagens do sensor TM a bordo do satélite Landsat-5. Depois dessa etapa, os dados da árvore foram utilizados para proceder a classificação digital da imagem no software ENVI© versão 4.3. As variáveis foram então inseridas no software e árvore executada, sendo que tomaram parte do

modelo as seguintes variáveis: bandas originais do sensor TM (1, 2, 3, 4, 5 e 7), Modelo Digital de Elevação (MDE) e o Índice de Vegetação da Diferença Normalizada (NDVI). Todas estas variáveis foram organizadas em um único arquivo de imagem.

Para avaliar o desempenho estatístico do método, no que tange a identificação da área plantada, foram realizadas análises comparativas entre as metodologias propostas com base nos valores efetivamente observados. Para o teste dos modelos, foram estimados os seguintes erros estatísticos: raiz do erro quadrático médio (REQM), erro absoluto médio (EAM) e o viés médio (VM) segundo WILLMOTT et al., (2005), as equações utilizadas na determinação dos erros estatísticos foram:

$$EAM = \frac{\sum_{i=1}^n |P_i - O_i|}{N} \quad (1)$$

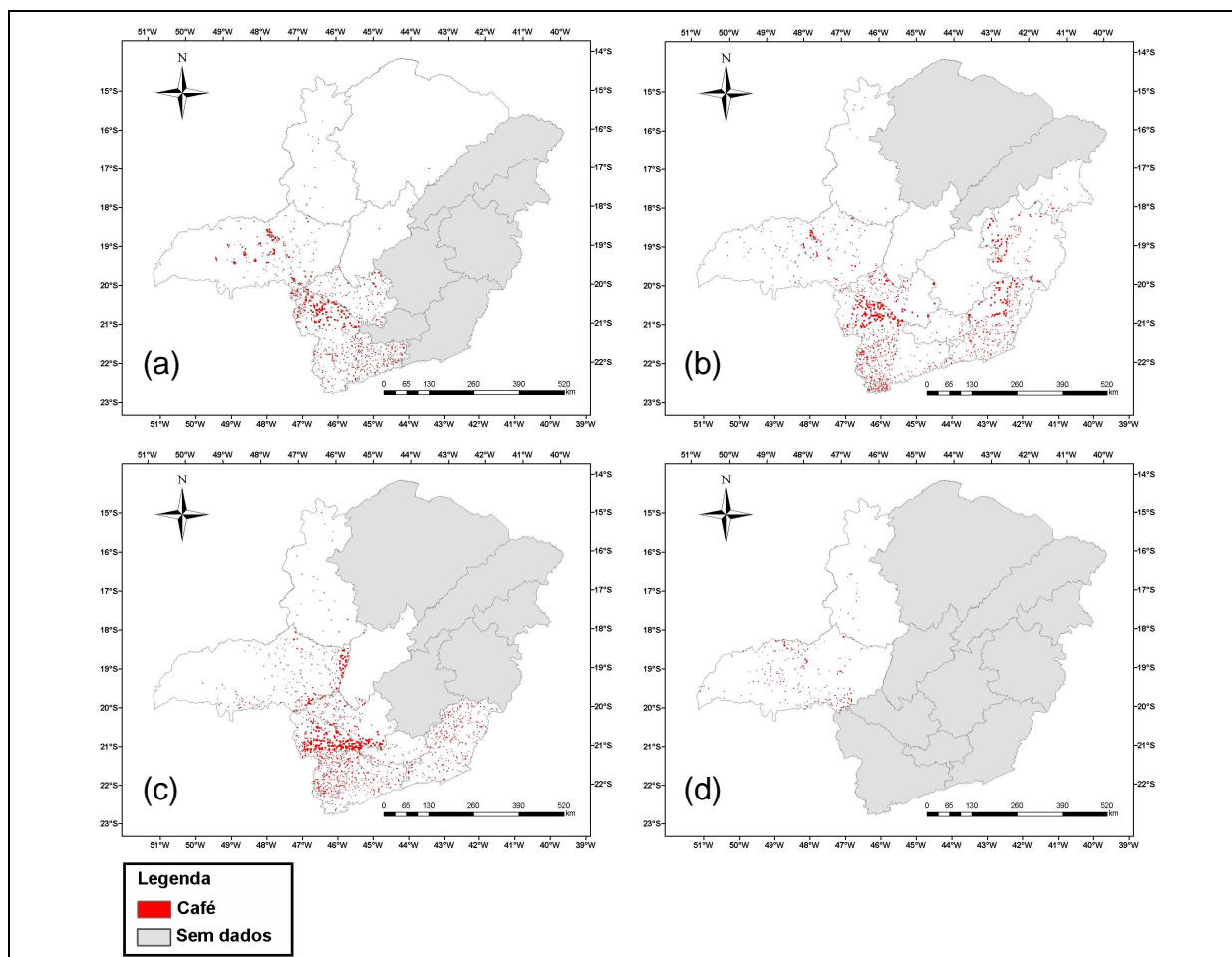
$$REQM = \left( \frac{\sum_{i=1}^n (P_i - O_i)^2}{N} \right)^{\frac{1}{2}} \quad (2)$$

$$VM = \frac{\sum_{i=1}^n (P_i - O_i)}{N} \quad (3)$$

em que,  $P_i$  é o  $i$ ésimo valor estimado,  $O_i$  é o  $i$ ésimo valor observado,  $\bar{O}$  é o valor médio observado,  $N$  é o número de pares de dados analisados.

## RESULTADOS E DISCUSSÃO

Na Figura 3a-d, observa-se os mapas gerados pela árvore de decisão da cultura do café para os anos de 1990, 1995, 2000 e 2005, nos respectivos anos, as áreas na cor cinza representam as regiões que foram desconsideradas.



**FIGURA 3.** Mapa de uso e ocupação do solo pela cultura do café, para os anos de 1990 (a), 1995 (b), 2000 (c) e 2005 (d) para o Estado de Minas Gerais, MG.

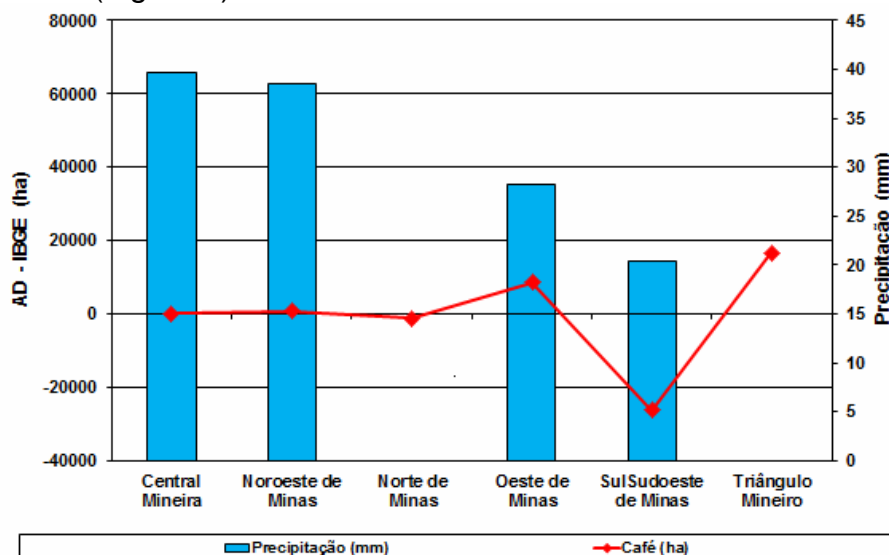
A Tabela 1 contém as informações estatísticas relacionadas à estimativa da área plantada por meio do classificador árvore de decisão e IBGE, onde verifica-se que para o ano de 1990 e 2005, a árvore de decisão subestimou os valores encontrados de área plantada de café, segundo a análise estatística do VM (-255,51 e -4.433,39 ha) e nos anos de 1995 e 2000 a árvore de decisão superestimou os valores de censo (4.596,96 e 13.689,14 ha) (Tabela 1).

PICOLI et al. (2009), ao utilizarem o uso de máscaras para o mapeamento das áreas plantadas com soja no Estado de São Paulo, encontraram resultados semelhantes, onde os valores estimados superaram os valores observados pelo IBGE, para as safras 2003/04 a 2006/07.

**TABELA 1.** Raiz do erro quadrático médio (REQM), erro absoluto médio (EAM), viés médio (VM), área total plantada, segundo dados do IBGE, e árvore de decisão (AD).

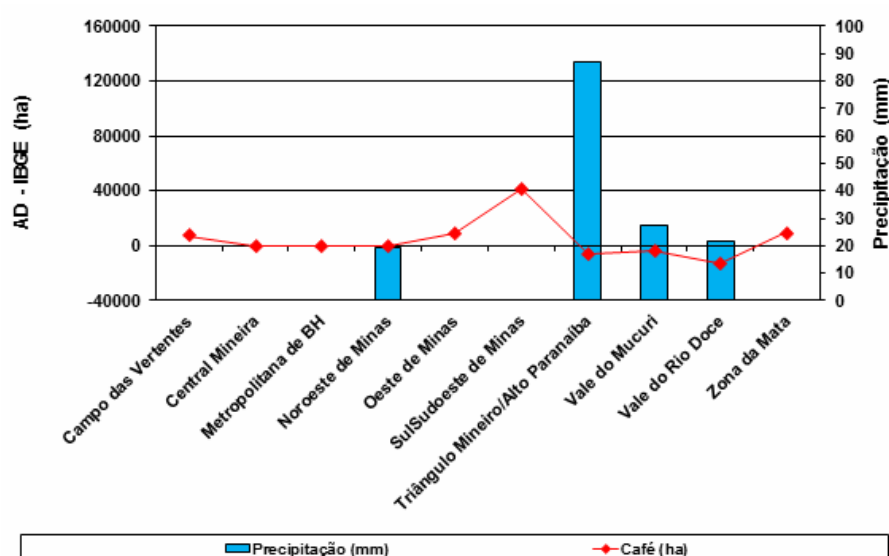
Cultura/Ano	REQM (há)	EAM (ha)	VM (ha)	IBGE (ha)	AD (ha)
Café/1990	13215,99	8899,90	-255,51	654293,00	652759,92
Café/1995	14714,71	8959,12	4596,96	819713,00	865672,56
Café/2000	28185,31	14041,00	13689,14	852822,00	948645,99
Café/2005	6816,63	5177,98	-4433,39	155788,00	146921,22

De acordo com os dados meteorológicos do Instituto Nacional de Meteorologia (INMET) e a Agência Nacional de Águas (ANA), as condições climáticas para o ano de 1990, para os 16 dias anteriores a passagem do satélite, para o Triângulo Mineiro, apresentavam-se com umidade relativa baixa e sem ocorrências de precipitações, o que resultaram em superestimativa com os dados observados para a cultura do café (Figura 4).



**FIGURA 4.** Valores de área plantada (AD) subtraídos dos valores observados (IBGE) versus precipitação para a cultura do café, para as mesorregiões para o ano de 1990.

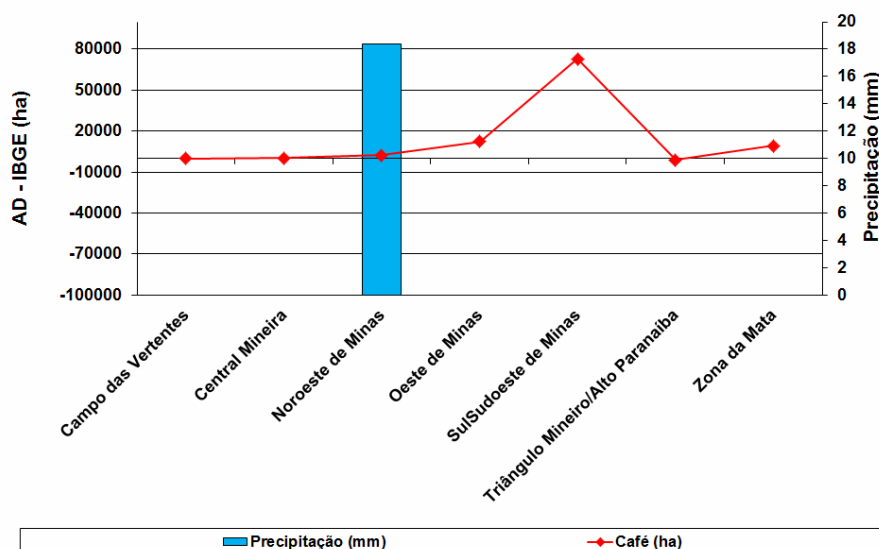
Em 1995, nos 16 dias anteriores da passagem do satélite, a mesorregião do Triângulo Mineiro foi a que obteve a maior precipitação (87,2 mm), a cultura do café teve seus valores de área subestimados quando comparado com os dados do IBGE (Figura 5).



**FIGURA 5.** Valores de área plantada (AD) subtraídos dos valores observados (IBGE) versus precipitação para a cultura do café, para as mesorregiões para o ano de 1995.



Na Figura 6, referente ao ano de 2000, a mesorregião Noroeste de Minas foi a única que, durante os 16 dias anteriores à passagem do satélite, ocorreu precipitação, dessa forma, houve superestimativa em relação aos valores observados.



**FIGURA 6.** Valores de área plantada (AD) subtraídos dos valores observados (IBGE) versus precipitação para a cultura do café, para as mesorregiões para o ano de 2000.

Apesar dos valores pluviométricos encontrados na mesorregião Noroeste de Minas no ano de 2005, a cultura café superestimou os valores observados de área plantada (Figura 7), assim, na região do Triângulo Mineiro ocorreu subestimativa de área plantada de café (Figura 7).

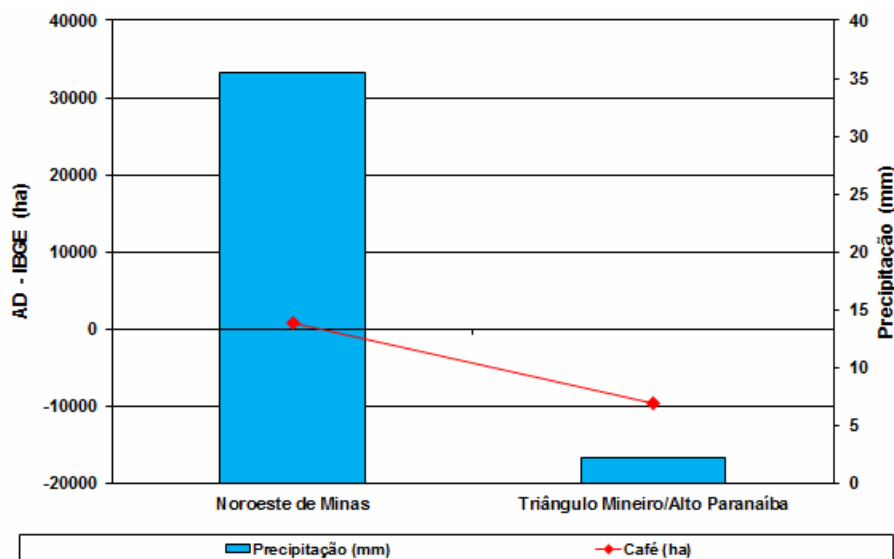
RIZZI et al. (2005), ao mapear as áreas plantadas com soja no estado do Rio Grande do Sul, observaram que no município de Palmitinho houve uma diferença de 1.083 ha a menos para a estimativa do Landsat (117 ha), ou seja, 9,2 vezes inferior à estimativa do IBGE. Exemplo mais drástico ainda é o de Santa Cruz do Sul, onde foram mapeados apenas 14 ha de soja nas imagens do Landsat, mas o IBGE informa que o município plantou 900 ha.

A discrepância entre os dados pode ser explicada, em parte, em função da metodologia de avaliação de safras que o IBGE utiliza, pois tal metodologia é subjetiva e consiste em fazer entrevistas com os produtores, cooperativas agrícolas a fim de se estimar os dados de área plantada. Também não são realizadas visitas a campo para mapear as propriedades e ter o conhecimento real das lavouras. O IBGE considera as localidades ou os produtores mais “representativos” para o levantamento de dados de produção e área, desta forma, existem possibilidades destes pequenos produtores ou pequenas propriedades serem desconsiderados no levantamento. Por outro lado, a discrepância sugere necessidade de validação em campo das informações levantadas no mapeamento de forma a inferir quantitativamente a confiabilidade das entrevistas a partir do parâmetro espacial. Isto tudo é decorrente da confiabilidade de aplicada nas informações fornecidas pelo



produtor/informante e por conseguinte, não há um parâmetro de confiabilidade nesta informação.

Com a disponibilidade crescente de satélites de sensoriamento remoto espera-se que num futuro próximo os órgãos oficiais responsáveis pelas estatísticas agrícolas se beneficiem das imagens de sensoriamento remoto para auxiliar no levantamento de informações objetivas sobre a área plantada com grandes culturas.



**FIGURA 7.** Valores de área plantada (AD) subtraídos dos valores observados (IBGE) versus precipitação para a cultura do café, para as mesorregiões para o ano de 2005.

## CONCLUSÃO

A classificação espectral por meio do algoritmo AD, mostrou que as imagens de sensoriamento remoto obtida pelo satélite Landsat-5, em quatro diferentes anos e específicas épocas, permitem mapear as áreas ocupadas com a cultura do café para o Estado de Minas Gerais.

Pode-se concluir nos anos estudados, que a classificação espectral do café foi sensível a eventos de precipitação, mostrando sempre uma tendência de redução de áreas em eventos desta variável. Esses resultados enfatizam a necessidade de se trabalhar com outras variáveis meteorológicas, além da precipitação.

## AGRADECIMENTOS

À CAPES e ao CNPq, pela bolsa de estudos concedida para a realização da pesquisa; e ao Programa de Pós-Graduação em Meteorologia Agrícola da Universidade Federal de Viçosa pelo uso do laboratório de Agrometeorologia.

## REFERÊNCIAS BIBLIOGRÁFICAS

ADAMI, M; MOREIRA, M.A.; RUDORFF, T.; FREITAS, C.C. Análise da eficiência dos estimadores de expansão direta e de regressão para áreas cultivadas com café, milho e soja no município de Cornélio Procopio, Estado do Paraná. **Agricultura em São Paulo**, v. 5, n. 2, p. 5-13, jul/dez 2004.

CARVALHO, L.G.; OLIVEIRA, M.S.; ALVES, M.C.; VIANELLO, R.L.; SEDIYAMA, G.C.; CASTRO NETO, P. DANTAS, A.A.A. Clima. In: Scolforo, J.R.S.; Carvalho, L.M.T.; Oliveira, A.D. **Zoneamento Ecológico-Econômico de Minas Gerais – componentes geofísico e biótico**. Lavras: Editora UFLA, 2008.

DELGADO, R. C.; SEDIYAMA, G. C.; COSTA, M. H.; SOARES, V. P.; ANDRADE, R. G. Classificação espectral de área plantada com a cultura da cana-de-açúcar por meio da árvore de decisão. **Revista Engenharia Agrícola**, Jaboticabal, v. 32, n.2, p.369-380, mar./abr. 2012.

EPIPHANIO, J.C.N.; LUIZ, A.J.B.; FORMAGGIO, A. R. Estimativa de áreas agrícolas municipais utilizando sistema de amostragem simples sobre imagens de satélite. **Bragantia**, v. 61, n. 2, p. 187-197, 2002.

Food And Agriculture Organization Of The United Nations (FAO). **Multiple frame agricultural surveys: current survey based on area and list sampling methods**. Roma, 1996. v. 1.

GONZALEZ-ALONSO, F.; SORIA, S. L.; GOZALO, J. M. C. Comparing two methodologies for crop area estimation in Spain using Landsat TM images and ground-gathered data. **Remote Sensing of Environment**, v.35, n.1, p 2 9-35, 1991.

PICOLI, M. C. A., ROCHA, J. V., ESQUERDO, J. C. D. M., LAMPARELLI, R. A. C. O uso de máscaras para seleção automática de áreas plantadas com soja no estado de São Paulo. **Anais do XIV Simpósio Brasileiro de Sensoriamento Remoto**, Natal, p. 333-338, 2009.

RIZZI, R.; RUDORFF, B. F. T. Estimativa da área de soja no Rio Grande do Sul por meio de imagens Landsat. **Revista Brasileira de Cartografia**. v. 57, n. 3, p. 226 – 234, 2005.

RIZZI, R.; RUDORFF, B.F.T. Imagens do sensor MODIS associadas a um modelo para estimar a produtividade de soja. **Pesquisa Agropecuária Brasileira**, v. 42, p.73-80, 2007.

WILLMOTT, C. J.; MATSUURA, K. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. **Climate Research**, v. 30, p. 79-82, 2005.